

A bibliometric reconstruction of research trails for qualitative investigations of scientific innovations

Gläser, Jochen; Laudel, Grit

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Gläser, J., & Laudel, G. (2015). A bibliometric reconstruction of research trails for qualitative investigations of scientific innovations. *Historical Social Research*, 40(3), 299-330. <https://doi.org/10.12759/hsr.40.2015.3.299-330>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

A Bibliometric Reconstruction of Research Trails for Qualitative Investigations of Scientific Innovations

Jochen Gläser & Grit Laudel*

Abstract: »Die bibliometrische Rekonstruktion von Forschungslinien für qualitative Untersuchungen wissenschaftlicher Innovationen«. Abrupt changes in research content are of interest to innovation research because many innovations in general and scientific innovations in particular emerge from such changes. However, investigations of innovations emerging from research processes face the problem that the initial change of direction in research by one or few researchers is an elusive phenomenon. The method presented in this article contributes to solving this problem by supporting the in-depth analysis of individual research biographies and of the emergence of new directions of research in these. The method employs bibliometric tools for a reconstruction of individual cognitive careers, embeds these reconstructions in qualitative studies of research biographies, and provides opportunities to link cognitive careers to the dynamics of scientific fields. As we will demonstrate, the method is generic in that it supports not only the investigation of scientific innovations but also, more generally, the identification of thematic change in individual cognitive careers. Two applications in qualitative research projects illustrate the potential of the method.

Keywords: Methods, innovation, scientific innovations, cognitive careers, interviews, bibliometrics, bibliographic coupling.

1. The Need for Innovation Research to Analyse Abrupt Change in Research Content

The method presented in this article supports the identification and analysis of abrupt change in research content. Such change is of interest to innovation research for two reasons. First, many innovations in society have their roots in scientific, social scientific or humanities research. This holds regardless of the model of innovation processes applied. Linear and co-evolutionary models of innovation processes have in common that they assign research (and abrupt changes therein) an important role in early stages of many innovations.

* Jochen Gläser, Center for Technology and Society, Technical University of Berlin, HBS1, Hardenbergstr. 16–18, 10623 Berlin, Germany; jochen.glaser@ztg.tu-berlin.de.
Grit Laudel, Department of Sociology, Technical University of Berlin, FH 9-1, Fraunhoferstr. 33–36, 10587 Berlin, Germany; grit.laudel@tu-berlin.de.

Second, research in the sciences, social sciences and humanities is itself driven by internal innovations that provide new research opportunities and open up new directions of research without ever ‘leaving’ the research fields. These specific innovations belong to the subject matter of innovation studies. Investigating them, however, poses specific problems because all research is ‘innovative’ in the sense that it creates new knowledge. For a concept of scientific innovation to make any sense, it must be used to characterise specific processes of knowledge production that can be distinguished from the incremental innovation inherent to all production of new knowledge. A possible conceptual solution to these problems, which we base this article on, is defining scientific innovations as research findings that affect the practices of a large number of researchers in one or more fields, i.e. their choices of problems, methods, or empirical objects (Laudel and Gläser 2014, 1207). Scientific and science-based innovations begin with one or few researchers significantly deviating from previous research practices, a deviation which subsequently diffuses in the community.

Investigations of science-based innovations and scientific innovations face the common problem that the innovations’ early phase, the first change or direction in research by one or few researchers, is quite an elusive phenomenon. The fluid and shifting knowledge landscapes in which research takes place and the esoteric nature of research processes create two methodological problems. The first is a ‘needle in a haystack’ problem – to identify specific kinds of change in a system consisting of nothing but change. The second problem is one of empirical access. Since an innovation can be clearly identified only after change occurred at the community level, its early phase can only be investigated retrospectively. Such an investigation usually depends on information that only the few researchers involved in the early phase can provide. Thus, changes in informants’ research biographies during the early phase must be identified and conditions for them must be retrospectively established. Cozzens’ (1989) study of the discovery of the opiate receptor illustrates that individual and collective retrospective rationalisations are likely to constitute a major methodological problem for such an investigation, especially if major scientific innovations are studied. No fewer than four laboratories claimed to have discovered the opiate receptor and could plausibly reconstruct their research in a way that supported their claim.

The aim of this paper is to present and discuss a method that contributes to solving these problems by supporting the in-depth analysis of individual research biographies and of the emergence of new directions of research in these biographies.¹ The method employs bibliometric tools for a reconstruction of individual cognitive careers in interviews, embeds these reconstructions in

¹ This method has evolved over several years, and a first description has been presented to bibliometricians (Gläser and Laudel 2009). We gratefully acknowledge its application in several projects and helpful comments by Enno Aljets, Julien Barrier, Jana Bielick, Elias Håkansson, Robert Jungmann, Stefan Lange, Eric Lettkemann, Sarojini Martin and Richard Woolley.

qualitative studies of individual research biographies, and provides opportunities to link cognitive careers to the dynamics of scientific fields. As we will demonstrate, the method is generic in that it supports not only the investigation of scientific innovations but also, more generally, the identification of thematic change in individual research biographies. We begin by identifying the diachronic structures in the course of a researcher's work that constitute either the content of an innovation or at least an important condition for its emergence (2). We then discuss opportunities to reconstruct these structures from a researcher's oeuvre, and describe our method for doing so (3). Two applications in qualitative research projects illustrate the potential of the method (4). As a conclusion, we discuss limitations of the method and opportunities for its improvement (5).

2. Diachronic Structures in Individual Research Biographies

The abrupt changes in research content that mark the emergence of scientific innovations can be identified against the continuity of diachronic knowledge structures. By diachronic knowledge structures we mean networks of research processes conducted at different times, which are connected through time by output-input relationships, i.e. because they build on each other. Research in the sciences, social sciences and humanities differs from much other organised work because it is embedded in such diachronic structures, which it simultaneously extends.

The diachronic knowledge structure that is central to the emergence of scientific innovations emerges and operates on the individual level. It evolves in the course of an individual researcher's cognitive career because each research process uses and contributes to the researcher's evolving scientific knowledge.² Each project contributes to that knowledge, which in turn is used in current and future research. A researcher's problem solving processes are interlinked through this use of previously produced knowledge in subsequent research. They form 'research trails,' i.e. sequences of thematically interconnected projects in which findings from earlier projects serve as input in later projects (Chubin and Connolly 1982). These trails not always take the simple linear form considered by Chubin and Connolly. Researchers may simultaneously

² We developed the concept 'cognitive career' as part of a model of academic careers. This model builds on two suggestions by Barley, namely to revive ideas of the 'Chicago School' for the research on careers (Barley 1989) and to 'bring work back in' the analysis of organisations (Barley and Kunda 2001). Combining both suggestions, we conceptualise academic careers as consisting of three analytically separable but closely interlinked careers, namely an organizational career, a status career in the scientific community, and a cognitive career consisting of successive stages of knowledge production building on each other (Laudel and Gläser 2008, 2011).

work on several different topics, i.e. have several parallel research trails (see e.g. Zuckerman and Cole 1994, 398-9; Gläser and Laudel 2007, 143). A research trail can branch out into several new trails, which are relatively independent of each other and are followed by the researcher in parallel. Research trails may also end if the researcher loses interest or does not find funding for continuing them.

A similar phenomenon occurs on the level of research groups or collaboration networks. Depending on the stability of such groups or networks, they, too, may construct specific bodies of knowledge to which all projects by group or network members contribute, and which evolve in a group's 'cognitive career.' The structure of a group's cognitive career (the network of its interlinked research trails) is mainly shaped by permanent and long-term group members. Nevertheless, all members of a group or network including transient members contribute to the group's or network's cognitive career.³

Through these diachronic structures, previous research influences the choice of new research problems, as current research will – through modifying the structures – influence future choices. The evolving bodies of knowledge constitute important conditions of action for researchers.

We can now further specify what we mean by abrupt changes in the content of research. The events marking the emergence of scientific innovations are likely to trigger new research trails on individual and group levels. Identifying these 'birth events' of new research trails means identifying situations in which an innovation might have emerged, and therefore enables the in-depth investigation of these situations.

3. Bibliometric Methods for the Analysis of Cognitive Careers

The diachronic structures of interest are structures of research content, which creates a major methodological challenge because the analyst rarely understands this research content. All approaches to the analysis of cognitive careers therefore depend on the analyst acquiring an understanding of the research at the level of an advanced lay person or, in the words of Collins and Evans (2002), the acquisition of "interactional expertise." If this can be achieved, it is possible to obtain information about the content of research from those who are

³ Although our empirical investigation of cognitive careers of research groups just began, we would like to venture the hypothesis that group-level cognitive careers are more than the sum of individual research trails because they reflect the collective nature of the research undertaken by the group. Thus, we assume group-level cognitive careers to be emergent phenomena. For example, they might address larger topics, which are only partially represented in each of the group members' cognitive careers.

most knowledgeable about it – the researchers themselves – without having to submit to the researcher’s own subjective theories about what happened and why (Laudel and Gläser 2007, 2012, 8-11).

Among the methods that can be used for analysing cognitive careers, bibliometric methods have the triple advantage of not requiring knowledge of research content, being ‘objective’ in the sense that they do not depend on a researcher’s interpretation of her cognitive career, and being based on a researcher’s ‘real-time’ decisions, i.e. on decisions made at the time the research took place. Bibliometric methods use properties of publications to identify the structure and dynamics of the knowledge contained in these publications. They thus exploit decisions made by researchers when they published their findings. ‘Objectivity’ means here that the analysis does not depend on ad-hoc interpretations of past actions, which avoids distortions by an informant’s retrospective rationalisations. This is why bibliometric methods are an excellent means for the triangulation of interview-based or ethnographic methods. However, they do not completely avoid the problem of analysing research content because the outcomes of bibliometric methods need to be interpreted and contextualised.

Taking into account these methodological challenges, we use bibliometric analyses of cognitive careers to support our qualitative, interview-based investigations. The purpose of our application of bibliometric analyses is to identify research trails in the interviewee’s cognitive career and to create a visual representation of these trails that can be used to explore researchers’ cognitive careers in interviews. This approach supports the interviewer’s acquisition of ‘interactional expertise,’ the discussion of research content in the interview, and the subsequent data analysis.

3.1 Bibliometric Approaches to the Reconstruction of Research Trails from a Researcher’s Publications

Diachronic knowledge structures are partially represented by sets of thematically connected publications. Although researchers and thus communities hold informal knowledge as well, the published (formal) knowledge consists of those knowledge claims researchers want their community to know about and use. This publicly available knowledge can be understood as the core of a community’s knowledge, and can be unobtrusively studied by bibliometrics. The investigation of knowledge structures by bibliometrics is based on the assumption that thematic links between research projects are reflected in similarities between publications resulting from these projects.

The identification of thematic structures in sets of publications has been one of the central concerns of bibliometrics for a long time. The search for methods has focused on the level of scientific communities (fields) and their topics (Small and Griffith 1974; Van Raan 1997, 215; Van den Besselaar and Heimeriks 2006). More recent bibliometric research was interested in the identification of ‘hot

topics' (Tseng et al. 2009) or 'emerging topics' (Glänzel and Thijs 2012), and in discovering diachronic structures at the community level by tracking topics over time (Small 2006; Mark, Roberts and Natali 2010).

Only few attempts have been made to investigate cognitive careers with bibliometric methods. The oeuvres of individual scientists have been analysed for a variety of purposes including

- mere description (Kalyane and Munnolli 1995),
- the exploration of methodological issues (see e.g. White 2000, 2001; Horlings and Gurney 2013),
- the creation of quantitative profiles of individual researchers (e.g. describing the evolution of numbers of publications, citations and co-authors over time, Zhang and Glänzel 2012); and
- analyses of a researcher's oeuvre aimed at the identification of field mobility (Hellsten et al. 2007).

The study by Hellsten et al. is the only one that applies bibliometric methods in an analysis of research content. They used an analysis of self-citations for identifying topics in the research biography of one researcher. The other authors conducted formal analyses of citation behaviour and citedness or attempted to advance methodologies. Horlings and Gurney interviewed some of the researchers whose oeuvres they analysed in order to validate their method.

Thus, in spite of some recent attempts to advance the methods for studying cognitive careers, the analysis of individual research biographies has not yet enjoyed much attention, probably because the uses of such analyses lie outside the field of bibliometrics.

The methodological suggestion that is closest to ours is that by Horlings and Gurney (2013). The authors combined two measures for the similarity of publications, namely bibliographic coupling and lexical coupling. Bibliographic coupling occurs if the same publication occurs in the reference list of two other publications. The shared reference is said to bibliographically couple the two publications that cite it. The number of references shared by two publications can be interpreted as an indicator of thematic similarity. Similarly, lexical coupling occurs if two publications use the same words or terms. Horlings and Gurney used shared words in titles of papers as a second measure of thematic similarity. The authors applied the Louvain algorithm (Blondel et al. 2008) for identifying topics in researchers' biographies.

We followed our own prior work (Gläser and Laudel 2009) rather than the suggestion by Horlings and Gurney (2013) because the latter's proposal raises some doubts. Although the authors obtained confirmation of the clusters they produced from some researchers whose oeuvre they investigated, the method they used is likely to produce artefacts for two reasons. First, the number of papers sharing both title words and references is relatively low (Van den Besselaar and Heimeriks 2006), which may lead to small topics not being identified and connections between papers belonging to the same topic not being found. Sec-

ond, the Louvain algorithm used by the authors, which was developed by Blondel et al. (2008) as a method for the fast identification of clusters in extremely large networks, maximises the modularity of the clustering solution. It is not clear at all why a modularity-maximising algorithm should be applied to the analysis of a scientist's cognitive career because overlaps of trails in these careers are both likely and theoretically interesting. Horlings and Gurney found an average number of research trails per physicist of more than ten and a range of four to 33, numbers that are highly unlikely and probably an artifact of the strong criterion for thematic links and of modularity maximisation. Therefore, we are reluctant to recommend this method until further tests of its validity and reliability have been conducted.

3.2 Reconstructing and Visualizing Individual Research Trails

We use bibliographic coupling for establishing paper similarities. Bibliographic coupling is today considered to be one of the best indicators of thematic similarity (e.g. Ahlgren and Jarneving 2008, 274-5). Compared to self-citations, which either link or do not link publications, bibliographic coupling provides links whose strength varies depending on the number of shared references and the length of the two reference lists.⁴ The relative strength of thematic connections is determined using Salton's Cosine, which is calculated as the ratio of references shared by documents i and j , R_{ij} , normalized by square root of the product of the numbers of references in documents i and j , R_i and R_j :

$$S = \frac{R_{ij}}{\sqrt{R_i * R_j}}$$

Salton's Cosine varies between zero for publications that do not share references and one for publications with identical reference lists.

The interviewee's publications are obtained from the Web of Science database and imported in an Excel Spreadsheet. They are then analysed with VBA Macros.⁵ The Macros produce lists of

- all bibliographic coupling links between publications,
- all self-citation links between the publications,
- the most frequent co-authors of the interviewee, and
- a list of publications that can be matched to the graphical representation of research trails.

⁴ Other disadvantages of self-citations are the strong variation of this practice between fields (Aksnes 2003, 241; Glänzel, Thijs and Schlemmer 2004) and the tendency of self-citations to over-emphasize thematic links between publications because authors use their publications for alerting readers to their other work, which in some cases is only weakly related, or for establishing their authority in the field.

⁵ Visual Basic for Application, the programming language embedded in Microsoft Office.

The macros also produce input files for the network analysis program Pajek, namely a network file and a vector file. The network file contains a list of vertices (representing publications) and their coordinates as well as a list of lines between vertices (representing bibliographic coupling) and information about their strength. X coordinates of vertices are derived from the publication year, while y coordinates are varied in order to evenly distribute vertices on the y-axis for the sake of better readability of the initial network graph. The vector file contains information about the size of vertices, which is derived from the number of citations a publication received.⁶ Pajek is used to identify separate clusters and to draw a graph showing these clusters. The graph is then exported to a drawing programme for further refinement.

The information produced by this procedure is used to prepare an interview, to create a graphical representation of the interviewee's research trails for discussion in the interview and to support the analysis of the interview. The graphical representation of research trails is based on a threshold for the strength of bibliographical coupling. Only publications that are bibliographically coupled with a strength above that threshold are considered to belong to the same research trail, which means that only lines between these publications are included in the input file for Pajek. The thresholds are not only specific to the investigated field but also specific to the individual cognitive career. They depend on the length of reference lists and on the extent to which researchers re-use references in articles that belong to the same or different research trails. Although these referencing practices are not completely arbitrary, they also depend on personal styles, which makes finding a suitable threshold a matter of trial and error. In most cases, experiments with thresholds lead to a graphical representation that has a recognisable structure, either in form of separate clusters of publications or by highlighting parts of the network that are not well connected to the network's main component.

We illustrate this method by reconstructing the cognitive career of the biologist Erica Larschan, a researcher who was portrayed in the newsletter *The Scientist* (Grens 2012). Reconstructing her cognitive career includes the following steps:

- 1) We searched in the internet for a list of publications, which we found on her website as part of her CV <<https://vivo.brown.edu/display/elarscha>>. The list was not up to date but it gave us sufficient information for an accurate search of her publications in the Web of Science. The CV provided us with information about her institutional affiliations.
- 2) We searched the Web of Science for Larschan's publications by using "Larschan E*" as search term in the 'author' field. A comparison of the publica-

⁶ Since the number of citations might become a sensitive issue for researchers who are not well cited, we scale the size of circles by always assigning a researcher's most highly cited publication the maximum circle size.

tions retrieved from the Web of Science to those in the publication list from Larschan's website showed that the Web of Science download included all publications from the publication list and some additional recent publications. We identified 17 publications (on August 13, 2014). The journal titles, publication titles and information on the author's institutional affiliation in the address field helped to make sure that the publications do indeed belong to this author.⁷

- 3) We downloaded the records from the Web of Science. The records included the following fields: authors, their addresses, title of the publication, journal title, volume, pages, keywords, document type (article, review, proceedings paper, etc.), the cited references, and number of citations received.
- 4) We conducted the bibliographic coupling analysis by processing the downloaded information with our VBA macros.⁸ The macros also produce a publication table (see Table 1).
- 5) Another VBA macro was used to create input files for the network analysis software Pajek. We chose a threshold value of 0.05 for Salton's Cosine, i.e. all links between publications below this value were excluded.⁹ The macro transforms information about the publications (publication year, number of citations received, and strength of bibliographical coupling) into information about the position and size of vertices as well as the existence and strength of lines between vertices, which is included in the two files described above.
- 6) We opened Pajek, loaded the input files, let the programme identify separate clusters, and constructed a picture with two separate clusters of publications.
- 7) The picture with the publication clusters was then exported from Pajek and imported in a graphics programme (we work with Microsoft Visio). The graphics programme is used to add information including the name of the scientist, a timeline of publication years, and a timeline of positions and organisations obtained from the CV. We named the clusters and included the titles of some particularly highly cited publications (see Figure 1).

Had we included Erica Larschan in one of our empirical investigations, we would have used the picture from Figure 1 and the publication list in our interview. Without the interview and a discussion of research trails with the researcher, the outcomes of the bibliometric analysis are of little use. They require interpretation, which can only be achieved by interacting with the researcher.

⁷ The problem of ambiguous author names (homonyms - different persons occurring in the Web of Science database with identical author names - and synonyms - one person occurring in the Web of Science database with more than one author name) is well-known in bibliometrics. Attempts to resolve this problem have been made by bibliometricians (e.g. Soler 2007; Tang and Walsh 2010) and recently also by the owner of the database, Thomson Reuters.

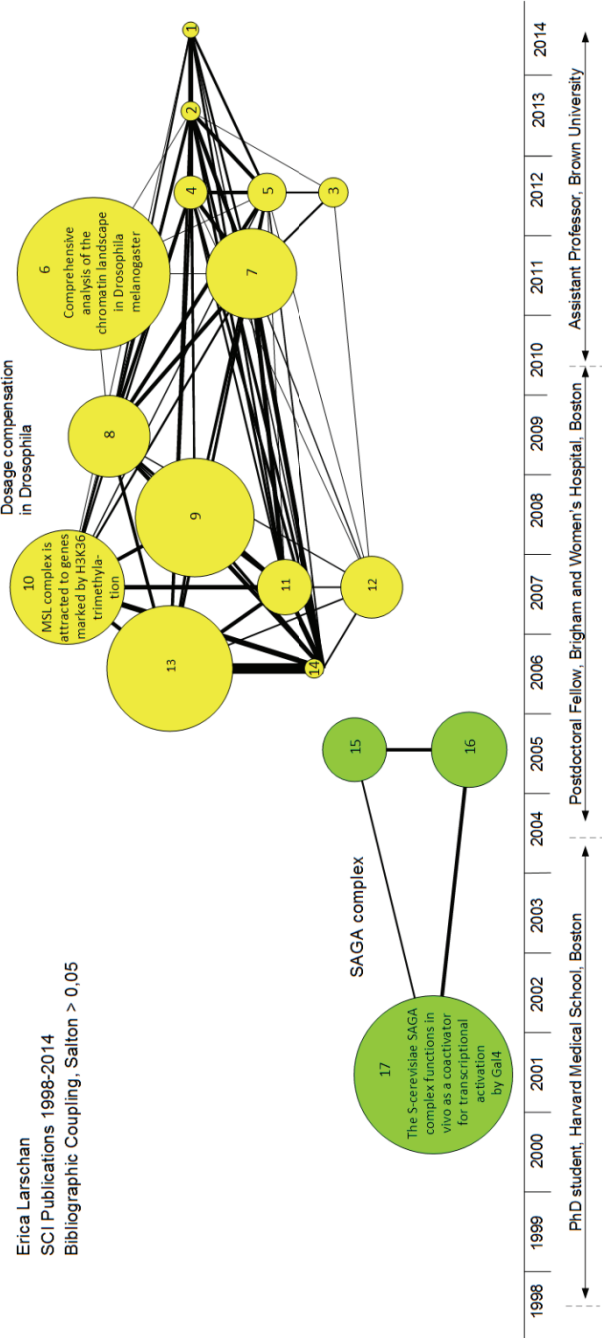
⁸ All technical details are described in a manual. The manual as well as the VBA Macros are available at <<http://www.laudel.info>>.

⁹ Since selecting a threshold is a matter of trial and error, steps 5 and 6 are usually repeated several times with varying thresholds in order to assess the stability of clusters.

Table 1: Publication List of Erica Larschan (Extract of the First and Last Three Publications)

| No | Authors | Year | Title | Journal | Publication Type | Keywords | Addresses | Times Cited |
|----|--|------|--|--|------------------|--|---|-------------|
| 1 | Chery, J; Larschan, E | 2014 | X-marks the spot: X-chromosome identification during dosage compensation | BIOCHIMICA ET BIOPHYSICA ACTA-GENE REGULATORY MECHANISMS | Review | Biochemistry & Molecular Biology; Biophysics / Dosage compensation; Drosophila; Mammals; C. elegans / LONG NONCODING RNAs; HISTONE H4-K16 ACETYLATION; MSI COMPLEX; DROSOPHILA-MELANOGASTER; ROX RNAs; CAENORHABDITIS-ELEGANS; SEX DETERMINATION; GENE-EXPRESSION; LINKED GENES; CHROMATIN | Brown Univ, Dept Biochem Mol Biol & Cellular Biol, Providence, RI 02912 USA | 0 |
| 15 | Prather, DA; Larschan, E; Winston, F | 2005 | Evidence that the elongation factor TFIIS plays a role in transcription initiation at GAL1 in Saccharomyces cerevisiae | MOLECULAR AND CELLULAR BIOLOGY | Article | Biochemistry & Molecular Biology; Cell Biology / RNA-POLYMERASE-II; PROMOTERS IN-VIVO; GENETIC INTERACTIONS; SAGA COMPLEX; FACTOR-SII; BINDING PROTEIN; ACTIVATOR; RECRUITMENT; MEDIATOR; YEAST | Harvard Univ, Sch Med, Dept Genet, Boston, MA 02115 USA | 32 |
| 16 | Larschan, E; Winston, F | 2005 | The Saccharomyces cerevisiae Srb8-Srb11 complex functions with the SAGA complex during Gal4-activated transcription | MOLECULAR AND CELLULAR BIOLOGY | Article | Biochemistry & Molecular Biology; Cell Biology / RNA-POLYMERASE-II; TATA-BINDING PROTEIN; CYCLIN-DEPENDENT KINASES; IN-VIVO TARGET; ACTIVATOR-TARGET; MEDIATOR COMPLEX; TERMINAL REPEAT; TBP BINDING; YEAST; RECRUITMENT | Harvard Univ, Sch Med, Dept Genet, Boston, MA 02115 USA | 45 |
| 17 | Larschan, E; Winston, F | 2001 | The S-cerevisiae SAGA complex functions in vivo as a coactivator for transcriptional activation by Gal4 | GENES & DEVELOPMENT | Article | Cell Biology; Developmental Biology; Genetics & Heredity / SAGA; Saccharomyces cerevisiae; Gal4; transcription; activation; TBP / TATA-BINDING PROTEIN; RNA-POLYMERASE-II; HISTONE ACETYLTRANSFERASE COMPLEXES; SACCHAROMYCES-CEREVISIAE; IN-VIVO; NUCLEOSOME ACETYLATION; YEAST TRANSCRIPTION; GLUCOSE REPRESSION; GENE DISRUPTION; DNA-BINDING | Harvard Univ, Sch Med, Dept Genet, Boston, MA 02115 USA | 202 |

Figure 1: Picture of Erica Larschan's Cognitive Career, which Appears to Consist of Two Research Trails



3.3 Use of Bibliometric Analyses of Cognitive Careers in Interviews

The use of graphical representations of cognitive careers contributes to the informational yield of interviews on several levels. First, it demonstrates the effort of the interviewer to understand the interviewee's work. The figure demonstrates "that you have done your homework, made an effort, and have not just come to pick their brain. You have gone as far as you can go with the available material and now you need some help" (Rubin and Rubin 1995, 198). We noticed that this often helped to build the necessary trust from the very start of the interview.

Second, the use of graphical representations contributes to creating a favourable atmosphere for the interview because the interviewee is confronted with a perspective on her own work she has not encountered before, i.e. the interviewer is not only asking for information but also providing some. Interviewees sometimes ask if they can keep the figures representing their work.

Third and most importantly, the representations of cognitive careers are used for 'graphic elicitation,' i.e. to prompt narratives about the content of research and to trigger memories.¹⁰ Interviews about events occurring several years ago are quite difficult because interviewees are often unable to recall events. The representation of a cognitive career, to which positions and grants held by the interviewee can be added from public sources, helps interviewees to reconstruct the evolution of their research.

The graphical representation of cognitive careers is introduced at the beginning of the interview, in most cases immediately after the introductory ('warming up') question. The interviewer briefly explains how the figure was constructed by

- referring to the Web of Science (a database scientists are familiar with) as the source of publications,
- naming but not explaining in detail the procedure by which clusters were produced (bibliographic coupling),
- explaining that the size of the circles indicates the numbers of citations a paper received, and

¹⁰ The use of other than verbal elicitation methods in interviews can be traced back at least to Collier (1957). Apparently, it emerged in anthropology with the use of photographs, which already has been extensively discussed (Harper 2002). The use of diagrams for "graphic elicitation" is a more recent trend (Crilly, Blackwell and Clarkson 2006; Umoquit et al. 2008, 2011). Sheridan, Chamberlain and Dupuis (2011) use graphic elicitation for "timelining," i.e. to support historical reconstructions by interviewees in a manner that is very similar to our approach (see also Umoquit et al. 2011, 3). Dempsey (2010) describes this use of visual material as "stimulated recall."

- linking the publications in the figure to a publication list by the numbers in the circles.¹¹

The discussion of the figure begins with specific but open questions, usually about the cluster representing the PhD project. It is of course impossible to discuss each publication in detail. The exploration of the figure therefore focuses on those clusters the interviewer is most interested in.

If separate clusters are visible in the figure, the interviewer asks whether these clusters represent thematic changes (the beginning of new research trails), and explores the reasons for these changes. In some cases, the figure will show only one cluster. If this is the case, the interviewer asks whether there are different topics hidden in this cluster. If the interviewee states that she has always worked on one topic, the reasons for continuity can be explored.

The figure might also show highly cited papers that are not connected to any cluster. If the interviewee is the first or last author of one of these publications, the interviewer asks about the topic of the publication and reasons why it might be disconnected in the figure.¹² Other disconnected publications are not usually discussed in the interview because they may represent contributions by the interviewee to the research of collaborators. These contributions are rarely of interest for an exploration of the interviewee's research trails.

In these discussions, interviewees occasionally correct the pictures by explaining that two trails that are separate in the figure are in fact thematically connected, or that one larger cluster contains two or more separate research trails. Such mismatches are inevitable for two reasons. First, the cluster analysis based on bibliographic coupling is an imperfect method. As all other features of publications that can be used for establishing thematic connections, shared references represent thematic similarities only to a limited extent, and arbitrary thresholds introduced to separate clusters can lead to misrepresentations. Second, the reconstruction of topics is necessarily ambiguous because it depends on the scientific perspective applied. Interviewees are aware of this ambiguity and sometimes mention it when explaining that treating two clusters of publications as separate or as belonging to the same topic depends on the scientific perspective applied to the picture. In any case, mismatches or ambiguities do not constitute a problem for the interview. The main function of these figures is to prompt narratives, which they also do when they make the interviewees correct the representation of their research trails (see on this topic also Crilly, Blackwell, and Clarkson 2006, 350). Corrections are recorded and used as additional information in the analysis of interviews.

¹¹ See Crilly, Blackwell and Clarkson (2006) on the introduction and explanation of diagrams in interview situations.

¹² First authorship usually indicates that the author conducted most of the experiments and drafted the article. Group leaders or heads of laboratories are usually last authors (see Laudel (2001) on roles in collaborative research and positions of authorship).

Thus, the graphical representation of an interviewee's research biography is used in interviews to identify changes in research trails and to discuss epistemic and other reasons for such changes. It makes it possible for the interviewer to discuss the content of research at a structural level, i.e. without having solely to rely on the interviewee's knowledge about research content. The research content can be explored at a depth both the interviewer and the interviewee are comfortable with without the investigation exclusively depending on the interviewee's epistemic authority concerning the content of his or her research.

In the analysis of interviews the bibliometric data are used to support the interpretation of interviews. In particular, the interviewee's narrative about her research biography can be linked to the research trails represented in publications. This enables a partial corroboration of statements about time lags between the intention to begin a new research line and the actual change of work, collaborations with other groups, and experimental difficulties that led to delayed publication. More importantly, the reconstruction of the cognitive career can be used to organise interview data for a 'process tracing' that supports the search for causal mechanisms (Gläser and Laudel 2013).

4. Applications

4.1. Conditions for Scientific Innovations

The first example stems from an internationally comparative analysis of the impact of changing authority relations on conditions for scientific innovations. We asked how the changing structures and processes of governance, through altering the relative authority of actors over the choice of research problems, modified the opportunities for researchers to develop innovations in their field. The project compared these conditions in four countries for one innovation each from the physical sciences, life sciences, social sciences, and humanities.¹³ The results of the project are described in several chapters in Whitley and Gläser (2014).

Our example is taken from the analysis of an innovation in physics, namely the experimental realisation of Bose-Einstein Condensation in cold atom gases. Bose-Einstein Condensation (BEC) is a state of matter that occurs when gases

¹³ For methodological reasons, our study did not concentrate on the actual innovation but rather included the early diffusion of the innovation in different national science systems. Innovations are likely to emerge in just one country, which is why an internationally comparative study of the emergence of innovations would have required finding similar innovations emerging in the countries under investigation at approximately the same time. This is close to impossible. Since the early development of an innovation poses similar problems as the actual invention in terms of contradicting the community's majority opinion, access to resources, and reputational risks, the early stages of diffusion were used and enabled comparisons of national science systems for the same innovation.

of atoms or subatomic particles are cooled to near absolute zero. This state had been theoretically predicted by Bose and Einstein in 1924. In 1995, two US groups produced the first BECs. It took more than two years before other research groups were able to replicate the experiments. When they succeeded, it soon turned out that BEC can be used for a wide range of fundamental research in several subfields of physics, and BEC research grew rapidly. Today more than hundred experimental groups and a multitude of theoretical groups worldwide work on BEC (for a detailed account of the case study see Laudel et al. 2014). However, until the early 2000s, researchers who wanted to produce a BEC faced significant technical and strategic uncertainties and needed considerable resources for the experiment to succeed.

We identified research groups from publications using the keywords ‘BEC’ or ‘Bose-Einstein Condensation,’ from internet searches of experimental physics groups at universities and from ‘snowballing,’ that is by asking interviewees about their national community. Interviews with researchers consisted of two main parts. In the first part, the interviewee’s research was discussed. Based on the bibliometric analysis and visual representation of the interviewee’s cognitive careers, we explored the development of the interviewee’s research since the PhD project with an emphasis on thematic changes and the reasons for them. In this part of the interview, developments in the interviewee’s national and international communities were also discussed. The second part of the interview explored conditions of research and the factors influencing them. Topics included the knowledge, personnel, equipment required to produce BECs, source of material support and opportunities as well as constraints provided by the interviewee’s academic posts.

In the following, we discuss the impact of authority relations on opportunities to innovate by comparing two researchers who wanted to begin research on BEC. One of them could immediately realise this intention, while the other’s move to BEC research was considerably delayed.

The interviews began by presenting a printout of the picture representing the interviewee’s cognitive career (Figure 2). For reasons of privacy protection we disguised the names of topics and publication titles as well as positions and locations.¹⁴

Q: Before we come to the Bose-Einstein condensation research, I would like to know something about your previous research, how it started. And what I tried to do is, I prepared a little picture where I put your – (“Oh, my goodness!”) – your Science Citation Index publications – (“Oh, how interesting!, wow!”) – on a time line – (“You put a lot of work in it.”). This is the list of

¹⁴ Interviews were conducted in English or German. German quotes were translated by us. Omissions and changes that are necessary for privacy protection are marked by [brackets], other cuts by “[...]”.

publications and the numbers you see here, 58 and so on, this is just the code for the publications.

A: And all these circles are different publications?

Q: Yes, different publications. And they are connected by a certain method we use where we look for joint references. In an area where we have no clue we can find out a bit about different topics.

A: Yes. Oh, that is interesting. You did a lot of work!

Q: Now, before you did Bose-Einstein condensation, you did earlier your PhD work, where you dealt with [a certain kind of] atoms.

A: [...] (*points at cluster III*) these are atoms [in a certain state]. And this is already the basis of what is coming there. Because that already sets a little bit the stage. As a physicist, how you develop what you do during your Ph.D. research is influencing you a lot what you do afterwards. And here (*pointing at cluster I*) you can already see that I was doing experiments on [B] atoms. [...] it is a kind of spectroscopy that I did.

The interviewee continued the narrative about the development of his topics, while constantly referring to the figure.

A: So after my PhD I had learned something about [a certain kind of] atoms and I wanted to do really more accurate spectroscopy. So I went to do a postdoc [abroad]. Where are my publications [from this time]? There are two of them, I hope.

Q: Let's see. (*searching through publication list*) [...] The 60, for example?

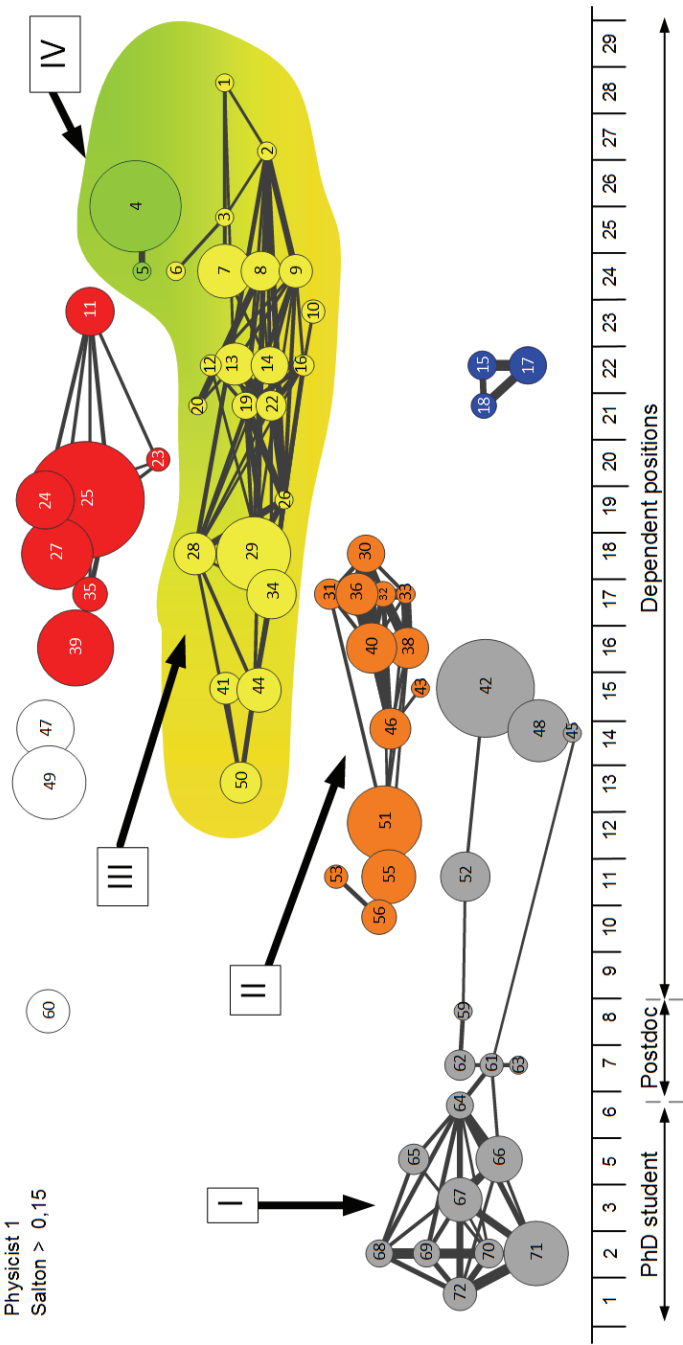
A: Yes. Here it is. That's what I did for my postdoc. There I learned more on highly accurate spectroscopy based on what I did here (*pointing at cluster I*).

[...]

And then I came back [...]. I got a position to start research. Then the whole field of laser cooling, which is the basis for Bose Einstein condensation had developed in the 80s. And my idea was to do laser cooling in [B] atoms because I knew [these atoms] from my PhD research. I can do spectroscopy in [B] atoms. I learned that [abroad] as well, that is very interesting. [...] And I had the idea, based on what I did [during my postdoc] and on my previous experience to set up [certain] experiments [...]. So I applied the laser cooling technique to do that kind of spectroscopy here (*pointing at cluster II*). It developed from here (*pointing at cluster I*).

From these sequences, we learn that the interviewee (as many other BEC researchers) had followed the international development and learned new techniques of laser cooling, which are important for producing BECs. The diagram not only triggered narratives about the start of new topics but also descriptions and explanations why certain topics ended.

Figure 2: The Cognitive Career of a Researcher who Moved to the Innovation BEC without Delay



A: That's why it is finished here (*pointing at cluster II*). I saw no future in that field. And similar things happened [elsewhere] there is not much continuing. It is more or less a completed field.

Q: Is it just saturated?

A: Yes it is just saturated. And that's what you have to do, you stop. [...] And I started to do laser cooling. And you can see that slowly the laser cooling field took over. And then I stopped this (*pointing at cluster II*) and I continued further on that (*pointing at cluster III*).

The interview continues with questions about international developments on Bose-Einstein condensation and about the start of the interviewee's own work on BEC. Inspired by the promising results and the eventual first experimental success of US groups, the researcher intended to move his research in this new direction and did so.

Cluster III in the picture contains both the use of laser cooling methods that was not yet directed at BEC and the subsequent work on BEC. Since the BEC research is the major focus of the interview, the interviewee was invited to 'zoom in' on cluster III and to provide more detail on this work.

Q: I noticed that you succeeded with the condensate in [year x]. And the experimental work is the yellow one (*cluster III*)?

A: Yes, it took [several] years.

Q: But you had continuously published on the topic. I wonder what these other publications are about, what could you publish?

A: Ah, yes. To make a Bose Einstein condensate you need to prove many things before that. For instance, for a Bose Einstein condensate you need a lot of atoms that are very dense. First you have to prove that you can have many atoms, that's the first publications. Not very cold, in a big magneto-optical trap. Then you want to show that the gas is stable. That is very crucial. And then you write some publications on that. And then you can do some studies of these ultracold gases, so you study the collision properties which are very important for Bose Einstein condensation. You can publish on that. ... All these studies on evaporative cooling – we did a real study how evaporative cooling works for [our kind of atom]. We published that. And spectroscopy. All these things that are related to the longer-term goal.

Q: Now, which one is actually the paper which reports the results about the Bose Einstein condensate? Would be about [year], I think. (*both search the publication list*)

A: Ah, this is it.

Q: Number 8. Ah, yes.

A: But then, this was important for this (*points at publication number 5 in the picture*). [...] and it still is a very important result because it opens new research directions.

This last remark of the interviewee revealed that the clusters III and IV (publication 5 and 4) are thematically connected, one being the production of a BEC (III), the other the use of a BEC in other research (IV).

The second part of the interview revealed important conditions that enabled the immediate move to BEC research. At this particular time, the researcher held positions on which he was formally dependent on his professor. The professor tolerated the plan to begin BEC research and granted the interviewee the use of infrastructure. The interviewee could mobilise the necessary additional funding for specific equipment and personnel first by ‘bootlegging’ money from other projects and later through grants specifically awarded for BEC research.

The interview with the second BEC researcher began again with the presentation of the diagram representing the cognitive career (Figure 3).

Q: Before I come to the [BEC] work which I’m most interested in, I would like to know something about your previous research. What I did is, I tried to reconstruct your publications and put them here on a time-scale. (“Wow! I have never seen this.”) It’s mainly the publications from the Web of Science. [...] And they get a link if they are thematically connected. We use a special method to find these connections via references. The size of the circles says something about citations. [...] My question is what had you been working on in your PhD?

A: My PhD work is this part of the figure here (*pointing at cluster I*). The title that you have here [...] is actually the title of my thesis. [...] This was in the field of atomic physics applying novel laser techniques to investigate new effects [...].

Q: There was nothing “cold” involved there?

A: No, this was room temperature or warmer.

The interviewee continues to describe his research trails.

A: And towards the end of my PhD I singled out a few groups where I would want to do postdoctoral work. I wanted to go [abroad], to see something of the world, but also scientifically to explore research at the highest levels on a worldwide scale. And I managed to get several offers. And the one I took – which has the orange blob here with all the citations (*pointing at cluster II*) – was [group X].

[...]

I started looking for postdocs [...]. To look ahead already a little bit, this explains this red area here (*pointing at cluster III*). This was work that I did in [City A] on a postdoc position.

This was the time when he actually considered beginning research on BEC but felt unable to compete with groups abroad, not least because he could not secure the considerable resources he would have needed:

A: If I would have had the opportunity to start a fully independent research group with the clear understanding that I would be completely on the same scale with these groups, things might have been different. [...] The investment necessary to do something close to identical to what was present at [the successful BEC groups in the US] would have required more than €500,000, per-

haps closer to €1 million, several junior people and several years of investment. This is not a scale where a junior person can start. I did not have the scientific muscle to bring about such a major effort.

He continued his career with dependent positions in research groups whose leaders had other interests than BEC research and whose infrastructure he could not utilize for his interest in BEC. Only when he could secure an independent position and obtained external research grants, he could begin the long-planned BEC work.

A: At that time I was really at the point where this [...] idea could take shape, start my own line of research, students that I guided from the start. So I started with one university-funded PhD student. I applied for funding for a second one. This was granted. Then one could really speak of a team and the setup was built. And this is the effort that is depicted here in yellow (*pointing at cluster IV*).

Thus, this interviewee's BEC research started with a delay of several years. It then took him several years to produce a BEC, apply it to answer an interesting question and to publish the results. This resulted in a publication gap.

Q: Yes, speaking about publications. I mean I am pretty much aware of the fact that Bose Einstein condensation is a several year effort, has been in many groups. And I always wondered: What do you do in the meantime with publications? Other groups, some of them could publish some more conceptual-theoretical stuff in the meantime until they achieved BEC. How was it in your case?

A: Yes. It was a conscious decision to focus on the experiment. I made this decision more or less consciously, also encouraged by my environment. The university did not ask me to publish papers.

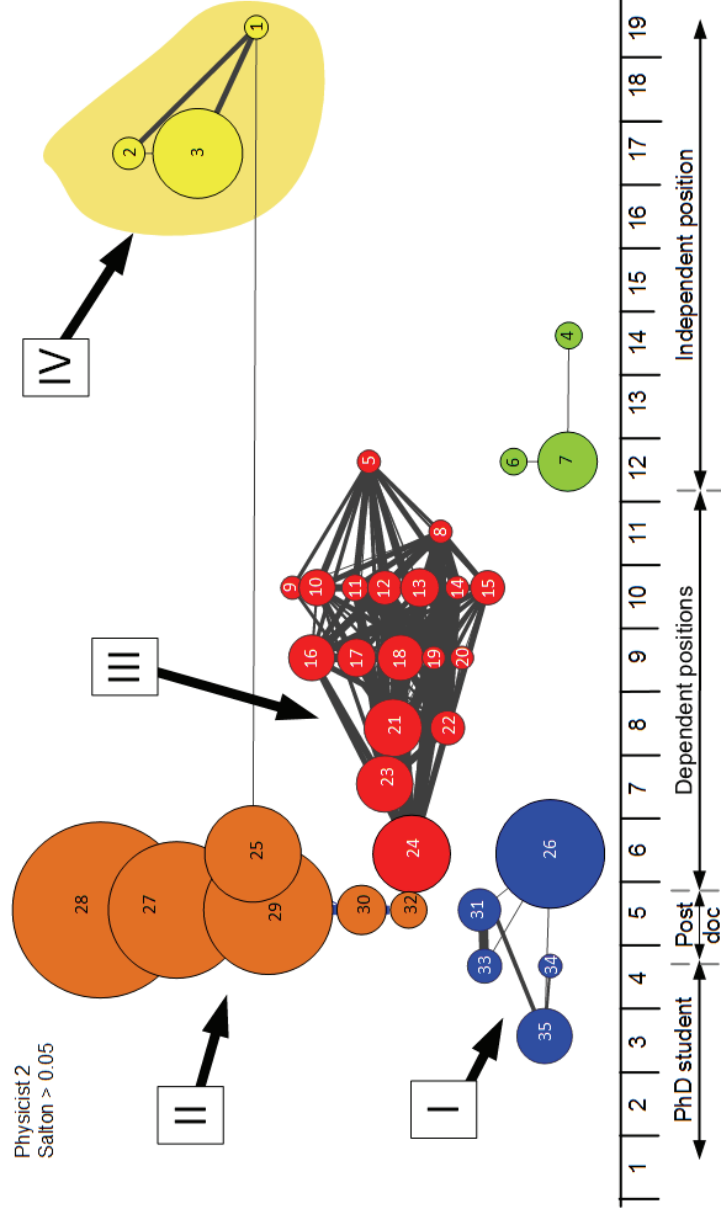
[...]

Yes, it was an all or nothing effort. I guess in my mind was the example that I had seen in the highly successful groups, back in '95, '96. Those groups also went through a clear publication dip of two or three years.

For this researcher, the conditions for moving to BEC research were unfavourable at first. He had not sufficient authority his own research goals or over the necessary infrastructure for pursuing them. He gained this authority only after obtaining an independent position. The university supported the move towards BEC by tolerating the lack of publication output for several years.

Taken together, interviews with BEC researchers enabled the identification of authoritative agencies with veto powers concerning the necessary conditions for moves to BEC research. Researchers who wanted to develop the innovation BEC needed to simultaneously control the infrastructure of a laboratory at the university, acquire external grants, have access to PhD students and work in an environment that tolerated publication delays of several years. Only if all these conditions were met the work with the scientific innovation was likely to be successful.

Figure 3: The Cognitive Career of a Researcher whose Move to the Innovation BEC was Delayed



The representations of research trails supported the identification of thematic changes and triggered narratives about the conditions under which they took place. At the same time, they also demonstrated another important point. Not all changes of research trails represent scientific innovations and not all are intentional. New research trails may also begin when dependent researchers move between research groups or in response to opportunities including funding opportunities, requests from collaborators, or access to methods and equipment.

4.2 The Emergence of Individual Research Programs in the Early Career Phase

Our second example uses data from an ongoing project that investigates the impact of early career researchers' conditions of work on the emergence of individual research programmes. Scientific communities expect researchers to become independent during the early career phase. In this phase, many early career researchers develop long-term interests that guide their selection of projects. These interests often take the form of individual research programmes. An individual research programme can be defined as a script for future research actions that contribute to the realisation of a larger research interest, i.e. to a goal that cannot be reached by a single project. Research programmes can lead to scientific innovations if their outcomes affect the scientific community by changing the practices of many of its members. However, most of them only lead to new research trails for the researcher who designs the programme without having any far-reaching community effects.

So far, the emergence of individual research programmes is poorly understood, which is why the project aims at ascertaining the conditions that shape the emergence of early career researchers' first research programmes. The project mainly draws on semi-structured interviews with German early career researchers in experimental physics, plant biology and history. Secondary data from previous projects on early career researchers which employed the same methodology are included (Laudel 2011, 2012).

The project's focus on thematic changes in a specific phase of the academic career suggests the use of bibliographic reconstructions of cognitive careers in interviews. In order to better support the aim of the project, the original technique was extended by introducing 'combined research trails.' With this extension, we responded to the observation that at least in the sciences, an early career researcher's independent research and first research programme are always prepared and triggered in a specific research environment that is not controlled by the researcher. As our preliminary results suggest, first individual research programmes usually emerge during the postdoctoral phase after the PhD, i.e. while early career researchers are members of a research group rather than already leading their own group. Consequently, the new research programmes emerge in a group that is currently following a different research programme, to which the author of the

new programme is also expected to contribute. This is why it is important to compare the research trails of early career researchers to those of their group leaders.

Our case example is Paul, a biologist who learned and used a specific method for his PhD project. After his PhD he took up a postdoctoral position abroad. In his new group he continued to use this method but applied it to different proteins and studied a different biological process.

A: At that point it was changing focus from looking at proteins involved in RNA processing to looking at proteins that are involved in another process, namely intracellular trafficking. But the techniques [...] were all things that I knew very well. So it was not so much a shift in methodology but a shift in the biological focus.

Paul's research trails show both the thematic continuity and the change he described in the interview (Figure 4). Cluster I represents his PhD work (which is confirmed by title and keywords and the co-authorship of the PhD supervisor). Clusters I and II are loosely connected (there is just one bibliographic link). Increasing the threshold for the bibliographic coupling to 0.13 separates the PhD work from cluster II. The postdoctoral project (cluster II) was defined by the group leader who had applied for the grant that funded Paul's postdoctoral position.

A: So [my group leader] wants to look at these protein machineries that drive the formation of these [cellular structures]. And the first question was what were the molecular structures of these machineries and what are the atomic details of how they interact with [regulatory complexes]. That is how I ended up in the field of protein trafficking and that is where I'm still working now. I have been continuing in that field since I started my postdoctoral work.

Paul explained in the interview that the general topic of his research evolved during his stay as a postdoc in the laboratory abroad, i.e. it was established by his group leader. This relationship changed during the postdoc phase when Paul's authority over defining research goals increased:

Q: Okay, yes. So, the project was really set for you?

A: Yes, it was. You got your diagram here. The initial work that I was involved in was looking at this [complex A] [...]. This probably happens in all the postdoctoral type situations – initially I started the project with a fairly defined goal, looking at this complex.

[...]

And from there the project evolved to look at different types of proteins that were involved in the formation of these particular [cellular structures]. These are these [G proteins]. And as this project evolved I was probably given a lot more leeway as to what I was most interested in. At the start we had a pretty specific goal, to look at [A] complex. That worked well. Then he said, these [G proteins] look quite interesting, why don't you have a look at those. So I did. But at that point I was given almost free range of deciding exactly what questions about these [G proteins] I found most interesting and how I go about doing that. That is actually an interesting graph here. So, and then as the post-

doctoral work leaped along, towards the end of my postdoc, after a few years I was almost given free range of deciding what particular questions I found most interesting, as long as they fell in this general field of trafficking, and in particular how it did fit with the other cell biologists in the department and what they were working on. And that's why I became interested in these particular proteins which is components of this [complex B]. *(points at the picture where cluster III starts)*

[...]

And that is now actually where the main projects are going on in my lab here.

[...]

The [complex B], I started working on it as a collaboration, as I said, with another group [at the university]. [...]. And what he found was that it seems to control a very interesting trafficking process in the cell. [...] We want to understand this fundamental process as to how this particular complex control the trafficking of these [proteins].

The research programme that Paul was pursuing at the time of the interview had emerged as a new research trail during his work as a postdoc. He continued this research trail and soon broadened it by studying the role of other proteins in the same biological process (cluster IV) after returning to his home country and becoming a group leader himself.

Paul claims that during his postdoctoral phase his group leader let him develop his independent research trail.

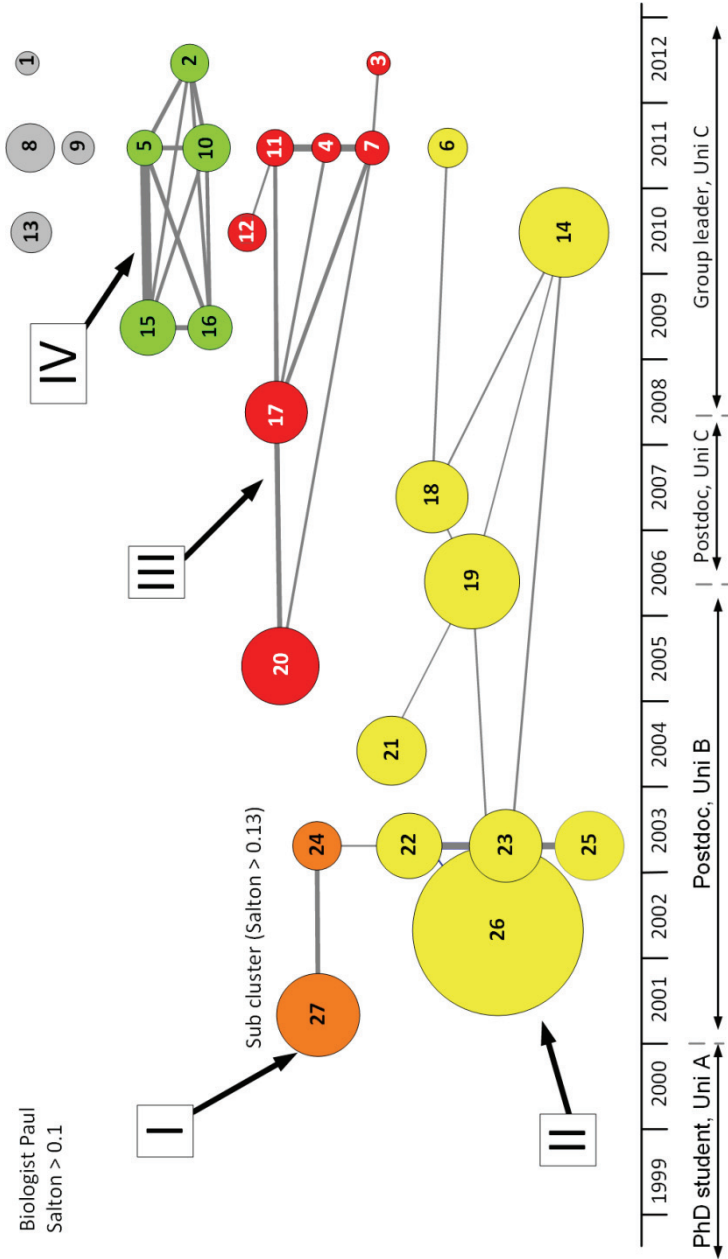
Q: Aha, ok. And this was already something that you developed independently from [your group leader].

A: Yes.

Q: This was one of my questions. I mean, [your group leader] did this trafficking work and he will continue to do it, but somehow you also have to make sure that you are not [...]

A: I was just very fortunate that [my group leader] was very supportive of doing that, developing my own career ... As you can see it, at the end I started to split my focus on the things that I was most interested in plus helping with other projects in the lab. And they allowed me then to develop a project which could be carried on as an independent researcher. And [my group leader] was very supportive of that and still is and still collaborates closely.

Figure 4: Cognitive Career of Biologist Paul

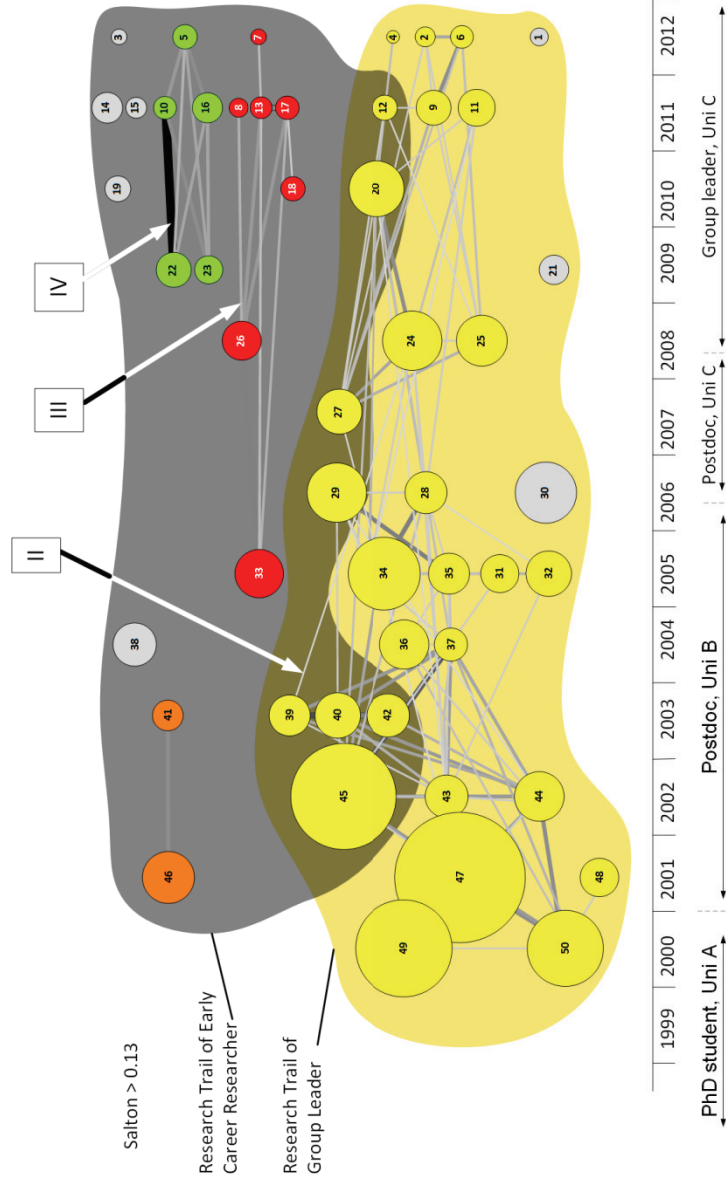


The interviewer wanted to independently test the reported scientific connections between Paul's work and his group leader's research. Therefore she additionally downloaded the group leader's publications from the Web of Science, beginning with publications from two years prior to Paul joining his lab. By choosing this time frame she could check whether Paul joined a research trail that already existed in the lab. Indeed, as Figure 5 shows, cluster II existed already before Paul came. The group leader continued to work on this topic after Paul left (lower area of the group leader's cluster). Paul had started to work on the group leader's topic. However, this topic largely faded out of his research. Only two recent publications (20, 12) are still strongly connected to the group leader's work.

The new topic Paul described in the interview (cluster III) has no bibliographic links to the group leader's work. The same holds for the second topic (cluster IV). Another indicator that these are indeed Paul's topics is that Paul has been occurring as last author or second-last author with increasing frequency. With cluster III, he started as first author (indicating that he did most of the experimental work), moving to the position of last author (indicating that he designed the research project) after he became a research group leader himself. The combined presentation of research trails showed that this early career researcher started a research programme that is indeed independent from his group leader's work.

We constructed such combined presentations of research trails for other researchers, too. In some cases, the separation of research trails could not be visualized although interviews and reviews of grant proposals confirmed the independence of former early career researcher and former group leader. This is due to the highly collaborative nature of research in this field. Collaborations that were started during a researcher's postdoctoral phase may be continued for a long time after the postdoc has left the lab and became a group leader himself. In all these cases, however, comparing the two representations to each other provides additional information for the analysis of the emergence of individual research programmes.

Figure 5: Combined Research Trails of Paul and the Group Leader of his Postdoctoral Period



5. Discussion and Conclusions

The method we presented here supports qualitative research on the emergence of subjectively or objectively new research topics in individual research biographies, a class of events we argued to include the ‘birth events’ of many scientific and science-based innovations. Although the method has some limitations that cannot be overcome, it supports interviews and observations through triggering narratives by ‘graphic elicitation’ or ‘stimulated recall’ and provides some opportunities to triangulate qualitative methods.

An important inherent limitation of the method that cannot be overcome is its equivocality. How topics are defined and delineated depends on the scientific perspective that is applied to the research. There is also the problem of bibliometric analyses working with properties of containers of knowledge rather than with the knowledge itself. The bibliometric reconstruction of topics depends on several arbitrary decisions, each of which may lead to different representations.

Equivocality means that one of several possible representations is used, which could constitute a problem in interviews because it is akin to asking leading questions. Fortunately, the interviewees we use this method with are trained in spotting errors, and arguing about wrong data is constitutive of their habitus. Researchers will object to rather than follow leads they consider to be wrong, which favours the application of our method because ‘incorrect’ representations are as good as ‘correct’ ones for triggering narratives.

Other limitations are more problematic. Since the method is based on the Web of Science database, its applicability depends on the coverage of publications of the field under investigation by this particular database. This means that the method as described in the paper is applicable only in most natural science and life science fields. It cannot be used for researchers from some fields in the sciences or from any field in the social sciences or humanities. For these fields, we currently use a workaround based on keywords obtained from titles. We are also currently experimenting with word-based network-analytic methods. However, this has proven more difficult than the use of bibliographic coupling and requires future work.¹⁵

Future work will also enable an extension of the method towards embedding individual researchers’ cognitive careers in research trails of their community. Since communities are much more fluid and perception-based social aggregates

¹⁵ We are grateful to Philip Roth for pointing out a possible technical problem of our method. Since more recent publications contain references that older publications cannot cite (because they were not yet published at that time), bibliographic coupling is likely to produce distortions by being biased against links spanning larger periods. Although our use of the method so far has not revealed distortions, systematic experiments are necessary to estimate the impact of time spans on bibliographic coupling.

than research groups, a different method of constructing their research trails is required. One possibility is the use of one researcher's research trails as 'seeds' for a 'greedy' algorithm, i.e. an algorithm that lets communities grow from seeds by stepwise adding the publications that are most similar to those already included.

These goals can best be reached by a collaboration between qualitative science studies and bibliometrics. Quantitative and qualitative methods and research designs were well integrated in the early days of science studies when, for example, the potential of the then-new Science Citation Index was explored (Cole and Cole 1967, 1968; Cole 1983; Hicks 1987; Pickering and Nadel 1987). In the 1980s, the two methodological approaches began to separate for several reasons, which were partly beyond the control of scholars in the field. Separate subfields emerged and developed a specific focus and accompanying methodologies. Science policy studies often rely on quantitative methods (both surveys and bibliometrics) because only these methods provide overviews of field-level or system-level dynamics. Not surprisingly, bibliometrics is focused on quantitative approaches, too. The sociology of science prefers qualitative methods because these methods can provide in-depth explanations of the mechanisms that produce changes in knowledge production and are thus ultimately responsible for field-level processes.

While it is difficult to integrate methods that differ in their underlying logics, conceptualisations, and empirical operationalisations of concepts, not doing so deprives research of important evidence and theoretical insights, thereby limiting its explanatory power.

References

- Ahlgren, Per, and Bo Jarneving. 2008. Bibliographic coupling, common abstract stems and clustering: A comparison of two document-document similarity approaches in the context of science mapping. *Scientometrics* 76: 273-90.
- Aksnes, Dag W. 2003. A macro study of self-citation. *Scientometrics* 56 (2): 235-46.
- Barley, Stephen R. 1989. Careers, identities, and institutions: The legacy of the Chicago School of Sociology. In *Handbook of Career Theory*, ed. Michael B. Arthur, Douglas T. Hall and Barbara S. Lawrence, 41-66. New York: Cambridge University Press.
- Barley, Stephen R., and Gideon Kunda. 2001. Bringing Work Back In. *Organization Science* 12 (1): 76-95.
- Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne LeFebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 10 (P10008): 1-12.
- Chubin, Daryl E., and Terence Connolly. 1982. Research Trails and Science Policies. In *Scientific Establishments and Hierarchies*, ed. Norbert Elias, Herminio Martins and Richard Whitley, 293-311. Dordrecht: Reidel.
- Cole, Stephen. 1983. The Hierarchy of the Sciences? *American Journal of Sociology* 89 (1): 111-39.

- Cole, Stephen, and Jonathan R. Cole. 1967. Scientific Output and Recognition, a Study in the Operation of the Reward System in Science. *American Sociological Review* 32 (3): 377-90.
- Cole, Stephen, and Jonathan R. Cole. 1968. Visibility and the Structural Bases of Awareness of Scientific Research. *American Sociological Review* 33 (3): 397-412.
- Collier, John. 1957. Photography in Anthropology: A Report on Two Experiments. *American Anthropologist* 59 (5): 843-59. doi: 10.1525/aa.1957.59.5.02a00100.
- Collins, Harry M., and Robert Evans. 2002. The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Science* 32 (2): 235-96.
- Cozzens, Susan E. 1989. *Social Control and Multiple Discovery in Science: The Opiate Receptor Case*. Albany: SUNY Press.
- Crilly, Nathan, Alan F. Blackwell, and P. John Clarkson. 2006. Graphic elicitation: using research diagrams as interview stimuli. *Qualitative Research* 6 (3): 341-66. doi: 10.1177/1468794106065007.
- Dempsey, Nicholas P. 2010. Stimulated Recall Interviews in Ethnography. *Qualitative Sociology* 33: 349-67.
- Glänzel, Wolfgang, and Bart Thijs. 2012. Using 'core documents' for detecting and labelling new emerging topics. *Scientometrics* 91 (2): 399-416. doi: 10.1007/s11192-011-0591-7.
- Glänzel, Wolfgang, Bart Thijs, and Báalazs Schlemmer. 2004. A bibliometric approach to the role of author self-citations on bibliometric macro indicators. *Scientometrics* 59 (1): 63-77.
- Gläser, Jochen, and Grit Laudel. 2007. Evaluation without Evaluators: The impact of funding formulae on Australian University Research. In *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, ed. Richard Whitley and Jochen Gläser, 127-51. Dordrecht: Springer.
- Gläser, Jochen, and Grit Laudel. 2009. Identifying individual research trails. In *Proceedings of the 12th International Society for Scientometrics and Informetrics (ISSI'09)*, ed. Birger Larsen and Jacqueline Leta, 841-5. Rio de Janeiro.
- Gläser, Jochen, and Grit Laudel. 2013. Life With and Without Coding: Two Methods for Early-Stage Data Analysis in Qualitative Research Aiming at Causal Explanations. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 14 (2), Art. 5 <<http://www.qualitative-research.net/index.php/fqs/article/view/1886>> (Accessed March 24, 2015).
- Grens, Kerry. 2012. Erica Larschan: Hitting Her Targets. *The Scientist* 26 (11): 55.
- Harper, Douglas. 2002. Talking about pictures: A case for photo elicitation. *Visual Studies* 17 (1): 13-26. doi: 10.1080/14725860220137345.
- Hellsten, Iina, Renaud Lambiotte, Andrea Scharnhorst, and Marcel Ausloos. 2007. Self-citations, co-authorships and keywords: A new approach to scientists' field mobility? *Scientometrics* 72 (3): 469-86. doi: 10.1007/s11192-007-1680-5.
- Hicks, Diana. 1987. Limitations of Co-Citation Analysis as a Tool for Science Policy. *Social Studies of Science* 17 (2): 295-316.
- Horlings, Edwin, and Thomas Gurney. 2013. Search strategies along the academic lifecycle. *Scientometrics* 94: 1137-60.
- Kalyane, Venkatrao. L., and Satish S. Munnolli. 1995. Scientometric Portrait of West, T. S. *Scientometrics* 33 (2): 233-56.
- Laudel, Grit. 2001. Collaboration, creativity and rewards: why and how scientists collaborate. *International Journal of Technology Management* 22 (7-8): 762-81.

- Laudel, Grit. 2011. *Exercising authority through academic careers – changes of authority relations in the Dutch university system*. 27th EGOS Colloquium, 7-9 July 2011, Gothenburg, Sweden.
- Laudel, Grit. 2012. *Doing something new in the Netherlands: The impact of research organisations and funding agencies on the start of new research lines*. 28th EGOS Colloquium, July 5-7, 2012, Helsinki, Finland.
- Laudel, Grit, and Jochen Gläser. 2007. Interviewing Scientists. *Science, Technology & Innovation Studies* 3: 91-111.
- Laudel, Grit, and Jochen Gläser. 2008. From apprentice to colleague: the metamorphosis of Early Career Researchers. *Higher Education* 55: 387-406.
- Laudel, Grit, and Jochen Gläser. 2011. Academic careers and how to find research excellence. *Plattform Forschungs- und Technologieevaluierung* (36): 3-13.
- Laudel, Grit, and Jochen Gläser. 2012. *The ERC's impact on the grantees' research and their careers*. EURECIA, Work package 4 summary report <<http://www.laudel.info/wp-content/uploads/2013/12/EURECIA-WP4-report-final-Jan2012.pdf>>.
- Laudel, Grit, and Jochen Gläser. 2014. Beyond breakthrough research: Epistemic properties of research and their consequences for research funding. *Research Policy* 43 (7): 1204-16.
- Laudel, Grit, Eric Lettkemann, Raphaël Ramuz, Linda Wedlin, and Richard Woolley. 2014. Cold Atoms – Hot Research: High Risks, High Rewards in Five Different Authority Structures. In *Organizational Transformation And Scientific Change: The Impact Of Institutional Restructuring On Universities And Intellectual Innovation*, ed. Richard Whitley and Jochen Gläser, 203-34. Bingley, UK: Emerald Group Publishing Limited.
- Mark, Herrera, David C. Roberts, and Gulbahce Natali. 2010. Mapping the Evolution of Scientific Fields. *PLoS ONE* 5 (5): e10355. doi: 10.1371/journal.pone.0010355.
- Pickering, Andy, and Edward Nadel. 1987. Charm Revisited: A Quantitative Analysis of the HEP Literature. *Social Studies of Science* 17: 87-113.
- Rubin, Herbert S., and Irene S. Rubin. 1995. *Qualitative Interviewing: The Art of Hearing Data*. London: Sage.
- Sheridan, Joanna, Kerry Chamberlain, and Ann Dupuis. 2011. Timelining: visualizing experience. *Qualitative Research* 11 (5): 552-69. doi: 10.1177/1468794111413235.
- Small, Henry. 2006. Tracking and predicting growth areas in science. *Scientometrics* 68 (3): 595-610. doi: 10.1007/s11192-006-0132-y.
- Small, Henry, and Belver C. Griffith. 1974. The Structure of Scientific Literatures I: Identifying and Graphing Specialities. *Science Studies* 4 (1): 17-40.
- Soler, José M. 2007. Separating the articles of authors with the same name. *Scientometrics* 72 (2): 281-90. doi: 10.1007/s11192-007-1730-z.
- Tang, Li, and John P. Walsh. 2010. Bibliometric fingerprints: name disambiguation based on approximate structure equivalence of cognitive maps. *Scientometrics* 84 (3): 763-84.
- Tseng, Yuen-Hsien, Yu-I. Lin, Yi-Yang Lee, Wen-Chi Hung, and Chun-Hsiang Lee. 2009. A comparison of methods for detecting hot topics. *Scientometrics* 81 (1): 73-90. doi: 10.1007/s11192-009-1885-x.
- Umoquit, Muriah J., Mark J. Dobrow, Louise Lemieux-Charles, Paul G. Ritvo, David R. Urbach, and Walter P. Wodchis. 2008. The efficiency and effectiveness of utilizing diagrams in interviews: an assessment of participatory diagramming and graphic elicitation. *BMC Medical Research Methodology* 8: 1-12.

- Umoquit, Muriah J., Peggy Tso, Helen E. D. Burchett, and Mark J. Dobrow. 2011. A multidisciplinary systematic review of the use of diagrams as a means of collecting data from research subjects: application, benefits and recommendations. *BMC Medical Research Methodology* 11: 1-10.
- Van den Besselaar, Peter, and Gaston Heimeriks. 2006. Mapping research topics using word-reference co-occurrences: A method and an exploratory case study. *Scientometrics* 68 (3): 377-93. doi: 10.1007/s11192-006-0118-9.
- Van Raan, A. F. J. 1997. Scientometrics: State-of-the-art. *Scientometrics* 38 (1): 205-18.
- White, Howard D. 2000. Toward Ego-Centered Citation Analysis. In *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, ed. Blaise Cronin and Helen Barsky Atkins, 475-96. Medford, NJ: Information Today, Inc.
- White, Howard D. 2001. Authors as Citers over Time. *Journal of the American Society for Information Science and Technology* 52 (2): 87-108.
- Whitley, Richard, and Jochen Gläser, eds. 2014. *Organizational Transformation and Scientific Change: The Impact of Institutional Restructuring on Universities and Intellectual Innovation*; vol. 42: *Research in the Sociology of Organizations*, ed. Michael Lounsbury. Bingley, UK: Emerald Group
- Zhang, Lin, and Wolfgang Glänzel. 2012. Where demographics meets scientometrics: towards a dynamic career analysis. *Scientometrics* 91 (2): 617-30. doi: 10.1007/s11192-011-0590-8.
- Zuckerman, Harriet, and Jonathan R. Cole. 1994. Research Strategies in Science: A Preliminary Inquiry. *Creativity Research Journal* 7 (3-4): 391-405.